

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
1 May 2003 (01.05.2003)

PCT

(10) International Publication Number
WO 03/036622 A2

(51) International Patent Classification⁷: **G10L 19/14**

(21) International Application Number: **PCT/EP02/11388**

(22) International Filing Date: 11 October 2002 (11.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
01250372.8 23 October 2001 (23.10.2001) EP
02090082.5 1 March 2002 (01.03.2002) EP

(71) Applicant (for all designated States except US): **THOMSON LICENSING S.A.** [FR/FR]; 46 Quai A. le Gallo, F-92100 Boulogne-Billancourt (FR).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SCHRÖDER, Ernst, F.** [DE/DE]; Pinkenburger Str. 25 D, 30655 Hannover (DE). **BÖHM, Johannes** [DE/DE]; Akazienweg 54, 37083 Göttingen (DE).

(74) Agent: **HARTNACK, Wolfgang**; Deutsche Thomson-Brandt GmbH, European Patent Operations, Karl-Wiechert-Allee 74, 30625 Hannover (DE).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

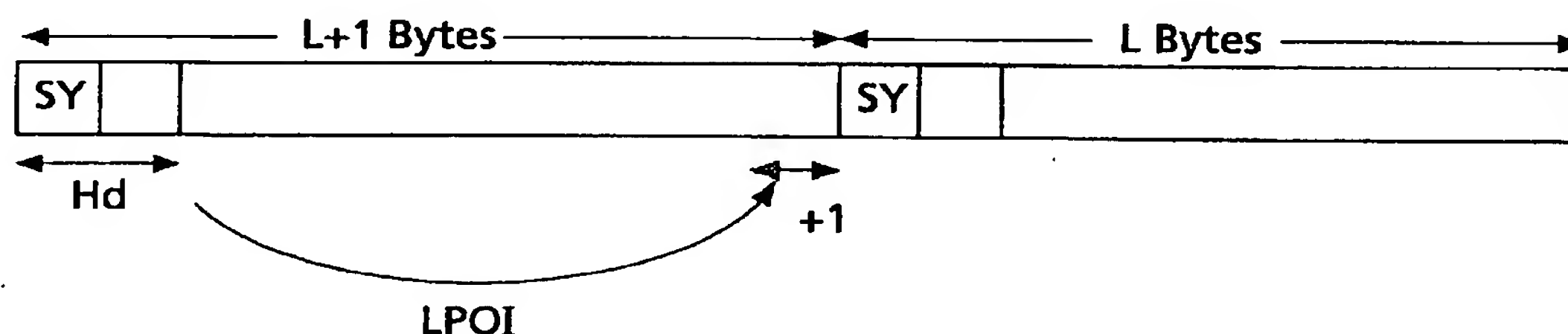
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD AND APPARATUS FOR DECODING A CODED DIGITAL AUDIO SIGNAL WHICH IS ARRANGED IN FRAMES CONTAINING HEADERS



(57) Abstract: With audio data reduction on the basis of ISO/IEC standard 11172-3, a frame length varying by 8 bits is used at a sampling frequency of 44.1 kHz in order to arrive, on average, at a particular fixed data rate. The lengthening of a data frame is signalled by a padding bit in the header of the frames. The invention dispenses with evaluation of the padding bit. Instead, the mean frame length L is calculated, L is rounded down to the next integer, for the subsequent frame it is first established whether the expected sync word for this frame appears, and, if this is so, this frame is decoded without taking into account the padding bit, but if the expected sync word for this frame does not appear, the decoding of the frame is started one 8-bit later without taking into account the padding bit.

Method and apparatus for decoding a coded digital audio signal which is arranged in frames containing headers

The invention relates to a method and an apparatus for
5 decoding a coded digital audio signal which is arranged
in frames containing headers.

Prior Art

10 When using audio data reduction on the basis of ISO/IEC
standards 11172-3 and 13818-3, a frame length varying
by 8 bits is used at a sampling frequency of 44.1 kHz
in order to arrive, on average, at a particular fixed
data rate (e.g. 128 000 bits/sec). The 'lengthening' of
15 a data frame is signalled by the "padding bit" in the
header of a frame. This method is described more
accurately in EP-A-0402973. The frames initially also
contain a sync word.

20 Invention

The evaluation of this padding bit in the decoder can
cause difficulties. By way of example, in highly
optimized decoders, the digital signal processors (DSP)
25 they contain require very sparing use of storage space.
Since, however, the header in a frame is read at the
start of decoding of the frame, but the value of the
padding bit is not needed until right at the end of
decoding of this frame, in a DSP implementation an
30 entire storage location (an integer value of, by way of
example, several bytes in length) is typically wasted
on merely storing the value of the padding bit.
It would be possible to achieve a reduction in the

required storage space by dispensing with the 'padding', i.e. the frame lengths would always be kept constant even at a sampling frequency of 44.1 kHz. However, a particular fixed data rate of, by way of example, 128 000 bits/sec is then no longer obtained, but rather a value which is 0.23% lower. A decoder which relies upon a constant frame length always being used even at a sampling frequency of 44.1 kHz would no longer be compatible with the aforementioned ISO/IEC standard, however.

The invention is based on the object of specifying a method which allows less storage space to be used but maintains the compatibility with the ISO/IEC standards 11172-3 and 13818-3 or with similar standards. This object is achieved by the method specified in Claim 1. A decoder using this method is specified in Claim 5.

In accordance with the invention, the data frames of varying length are evaluated on the basis of the respective length, but evaluation of the padding bit from the header is avoided. Since the value of the padding bit is normally used to ascertain the exact position of the start of the next frame, the invention involves ascertaining the start of the next frame in another way, namely by calculating a mean frame length and a rounding-down or rounding-up of this mean frame length to the closest integer byte values for the received frames.

The advantage is that the value of the padding bit does not need to be stored for the entire time taken for decoding a frame, and hence storage space can be saved

more frugally.

In principle, the inventive method relates to the decoding of a coded digital audio signal which is arranged in frames containing headers, where the header in a frame contains a respective information item regarding whether this frame has a standard length or a length which differs therefrom for some of the frames, and where the frames contain a respective sync word, having the following steps:

- the length-variation information regarding the respective frame length is not stored or evaluated;
- the approximate start of the next frame is determined using the following formula:

$$L = N * R / fs / SL,$$

where L is equal to the length of the frames, N is equal to the number of samples per frame, R is equal to the total data rate, fs is equal to the sampling frequency, SL is equal to the stipulated subunit for indicating the frame length;

- L is rounded down to the next integer of subunits SL;
- for the subsequent frame, it is first established whether the expected sync word for this frame appears;
- if the expected sync word for this frame appears, this subsequent frame is decoded without taking into account the length-variation information;
- if the expected sync word for this frame does not appear, the decoding of this subsequent frame is started one subunits later without taking into account the length-variation information.

In principle, the inventive apparatus relates to a decoder for decoding a coded digital audio signal which is arranged in frames containing headers, where the header in a frame contains a respective information item regarding whether this frame has a standard length or a length which differs therefrom for some of the frames, and the frames contain a respective sync word, where, for ascertaining the frame length, the length-variation information regarding the respective frame length is not stored or evaluated, and where the apparatus contains:

- means for decoding the audio signal;
- a frame-start estimator in which the approximate start of the next frame is determined using the following formula:

$$L = N * R / fs / SL,$$

where L is equal to the length of the frames, N is equal to the number of samples per frame, R is equal to the total data rate, fs is equal to the sampling frequency, SL is equal to the stipulated subunit, and in which L is rounded down to the next integer of subunits SL;

- a sync-word checker which, for the subsequent frame, first establishes whether the expected sync word for this frame appears, where, if the expected sync word for this frame appears, this subsequent frame is decoded in the decoding means without taking into account the length-variation information, and, if the expected sync word for this frame does not appear, the decoding of this subsequent frame is started in the decoding means one subunit SL later without taking into account the length-variation information.

Instead of evaluating the sync word, another known and expected data pattern can also be evaluated.

Drawings

5

Exemplary embodiments of the invention are described with reference to the drawings, in which:

Figure 1 shows two successive data frames having the same length;

10 Figure 2 shows two successive data frames having different lengths;

Figure 3 shows a decoder in accordance with the invention.

15 Exemplary embodiments

In data-reducing coding and decoding methods for audio signals, such as in ISO/IEC 11172-3 (MPEG audio), the coded audio signals are stored or transmitted in data
20 frames which respectively contain a fixed number N of audio samples, e.g. 1152 samples. The data frames have, in principle, a fixed length which is a multiple of a basic unit, which is called a 'slot' in ISO/IEC 11172-3 and has a length of 8 bits in the 'layer 2' and 'layer
25 3' variants.

In Figure 1, each of the successive frames of the same length of L bytes has a header Hd which contains a sync word SY. The size of the subunit SL is 1 byte = 8 bits
30 in this example.

If audio signals having sampling frequencies f_s of 32 000 Hz or 48 000 Hz are used, then the relationship

between the total data rate R (in bits/sec) and the frame length L (in slots) is as follows:

$$L = N * R / f_s / 8 \quad (1)$$

Example:

- 5 $N = 1152$ samples; $R = 128\,000$ bits/sec; $f_s = 48\,000$ Hz
gives $L = 384$ slots of 8 bits each.

If, however, a sampling frequency of 44 100 Hz is used, then non-integer values for L are produced in (1). In
10 this way, the start of the next frame is determined only approximately. Example:

$N = 1152$ samples; $R = 128\,000$ bits/sec; $f_s = 44\,100$ Hz
gives $L = 417.9591837$ slots of 8 bits each.

- 15 However, since a frame can only have an integer number of slots, a frame length which varies by 1 slot (= 8 bits) is used at a sampling frequency of 44.1 kHz in order to arrive, on average, at a particular fixed data rate (e.g. $R = 128\,000$ bits/sec) and is signalled, as
20 described above, using the padding bit in the header.

When the result from formula (1) is rounded down, the correct frame start is often obtained for a sampling frequency of 44.1 kHz, namely for those frames which have not been lengthened by 1 slot. Often, however, an
25 incorrect value is also obtained for the frame start. If the next frame starts to be decoded at this incorrect point, then an error is obtained, since the sync word to be expected at the start of the frame obviously does not appear.

- 30 Normally, decoders then switch to an error recovery mode and start a fresh complex search for a sync word. This typically produces a fault in the decoded output signal.

In figure 2, the first frame is one unit SL longer than the second frame, i.e. L+1 bytes. If decoding starts at the place pointed at by the pointer LPOI for the
5 calculated and rounded-down variable L, then no sync word is found at that place. For this reason, a check is carried out one unit SL further on to determine whether a sync word is present, and this sync word is found at that place.

10

The invention therefore proposes, when decoding encoded signals having the sampling frequency 44 100 Hz or 22 050 Hz:

- not storing or evaluating the padding bit;
- 15 - determining the approximate start of the next frame using the formula (1);
- rounding down the result from (1) to the next integer;
- for the subsequent frame, first establishing whether
20 the expected sync word or another known data pattern appears;
- if this is the case, decoding this subsequent frame without taking into account the padding bit;
- if this is not the case, starting the decoding of
25 this subsequent frame one slot later without taking into account the padding bit.

Figure 3 shows an inventive decoder which receives a coded audio signal EAS which is supplied to a bit
30 stream deformatter BSD. BSD interchanges corresponding data with a frame-start estimator FSE. The frame start address estimated therein or a corresponding pointer LPOI is used to establish, in a sync word checker SYCH,

whether there is a sync word at the appropriate point in the data stream. If this is so, a decoder stage DEC and/or the bit stream deformatter BSD receives the information which prompts the further processing or
5 decoding of the next data frame to start at that point. The audio signal decoded in the frequency domain is supplied by the decoder stage DEC to a windowing stage DW which multiplies portions of the audio signal using a synthesis filter, for example, converts them to the
10 time domain and outputs a decoded audio signal DAS.

The invention can also be used for related applications in which a non-integer result from (1) causes a variation in the frame length and said variation is
15 indicated using an information item similar to a 'padding bit'.

Patent Claims

1. Method for decoding (BSD, DEC, DW) a coded digital audio signal (EAS) which is arranged in frames containing headers (Hd), where the header in a frame contains a respective information item regarding whether this frame has a standard length (L) or a length (L+1) which differs therefrom for some of the frames, and where the frames contain a respective sync word (SY), characterized by the following steps:
- the length-variation information regarding the respective frame length is not stored or evaluated;
 - the approximate start of the next frame is determined (FSE) using the following formula:
$$L = N * R / fs / SL,$$
where L is equal to the length of the frames, N is equal to the number of samples per frame, R is equal to the total data rate, fs is equal to the sampling frequency, SL is equal to the stipulated subunit for indicating the frame length;
 - L is rounded down (FSE) to the next integer of subunits SL;
 - for the subsequent frame, it is first established (SYCH) whether the expected sync word for this frame appears;
 - if the expected sync word for this frame appears, this subsequent frame is decoded (DEC, DW) without taking into account the length-variation information;
 - if the expected sync word for this frame does not appear, the decoding (DEC, DW) of this subsequent frame is started one subunit later without taking

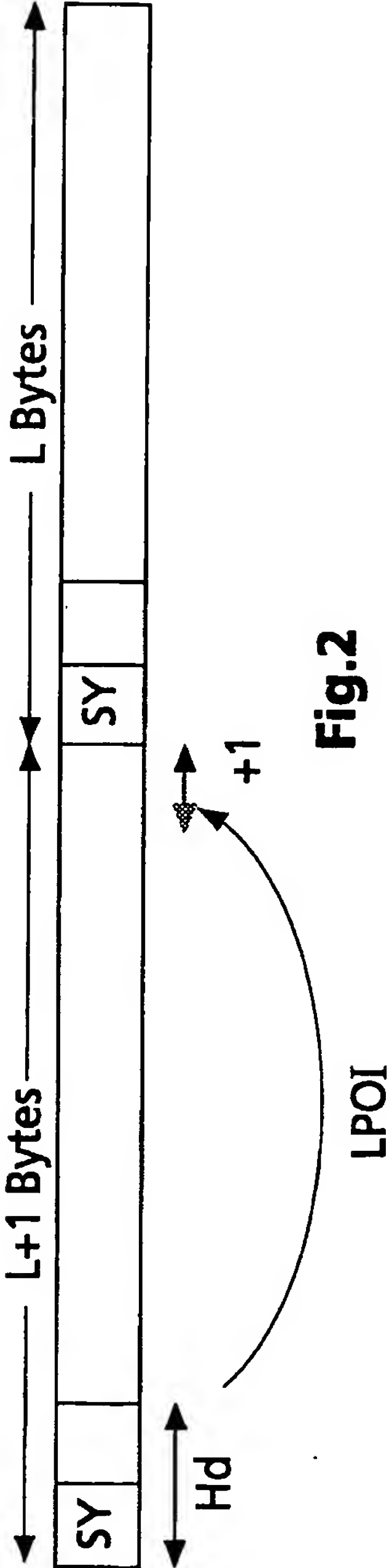
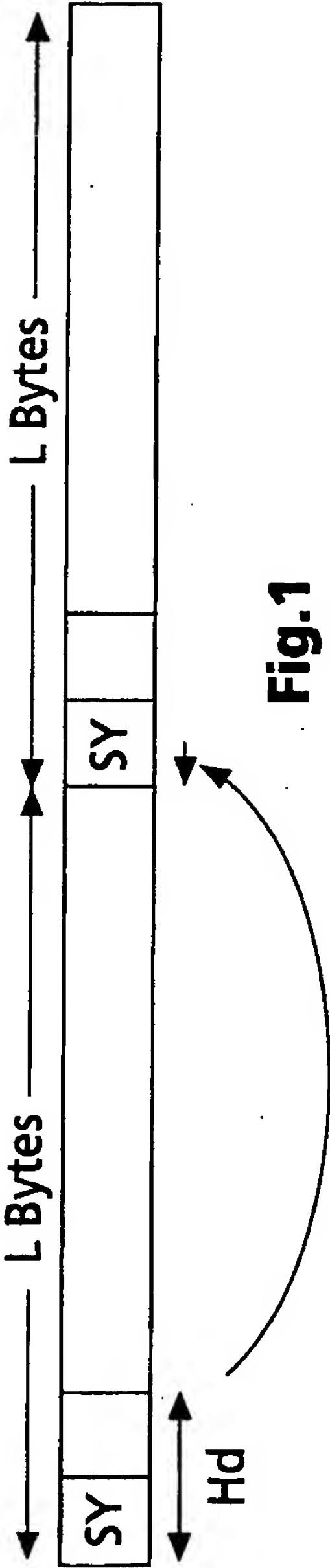
into account the length-variation information.

2. Method according to Claim 1, where the parameters
for calculating the formula for the approximate
5 frame start comprise known parameters in a
transmission system.
3. Method according to Claim 2, where at least one of
the parameters is transmitted in the header of
10 frames.
4. Method according to one of Claims 1 to 3, where,
instead of establishing whether an expected sync
word appears, it is established whether another
15 known pattern appears in the next frame.
5. Apparatus for decoding a coded digital audio signal
(EAS) which is arranged in frames containing headers
(Hd), where the header in a frame contains a
20 respective information item regarding whether this
frame has a standard length (L) or a length (L+1)
which differs therefrom for some of the frames, and
the frames contain a respective sync word (SY),
where, for ascertaining the frame length, the
25 length-variation information regarding the
respective frame length is not stored or evaluated,
and where the apparatus contains:
 - means (BSD, DEC, DW) for decoding the audio signal
(EAS);
 - 30 - a frame-start estimator (FSE) in which the
approximate start of the next frame is determined
using the following formula:

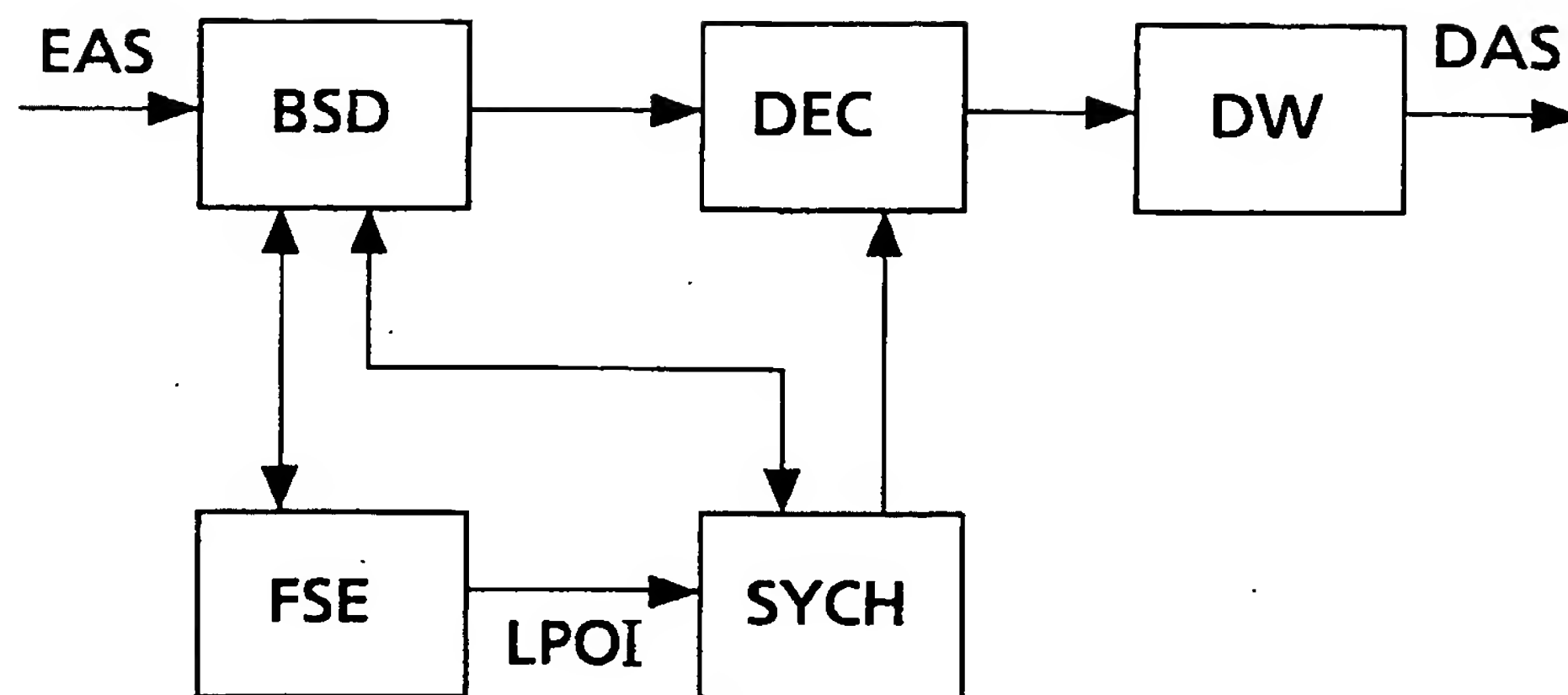
$$L = N * R / fs / SL,$$

where L is equal to the length of the frames, N is equal to the number of samples per frame, R is equal to the total data rate, f_s is equal to the sampling frequency, SL is equal to the stipulated subunit, and in which L is rounded down to the next integer of subunits SL ;

- a sync-word checker (SYCH) which, for the subsequent frame, first establishes whether the expected sync word for this frame appears, where, if the expected sync word for this frame appears, this subsequent frame is decoded in the decoding means without taking into account the length-variation information, and, if the expected sync word for this frame does not appear, the decoding of this subsequent frame is started in the decoding means one subunit SL later without taking into account the length-variation information.



2/2

**Fig.3**